

**FACULTY OF COMPUTING**

**SESSION (2023-2024)**

**Semester:1**

**Project Proposal**

**Course :** Big Data Management

**Course code :** MCSD1123

**Lecturer :** Prof. Madya. Ts. Dr. Mohd Shahizan bin Othman

| **Members** | **Matric No.** |
| --- | --- |
| MOHAMMED RAZA A. CHIDIMAR | MCS231004 |
| AYAZ RAHMAN BHUIYAN | MCS231023 |
| MUSAB IBNE AHMAD | MCS231017 |
| HUSSEIN YUSUF SHEIKH MOHAMED | MCS231024 |

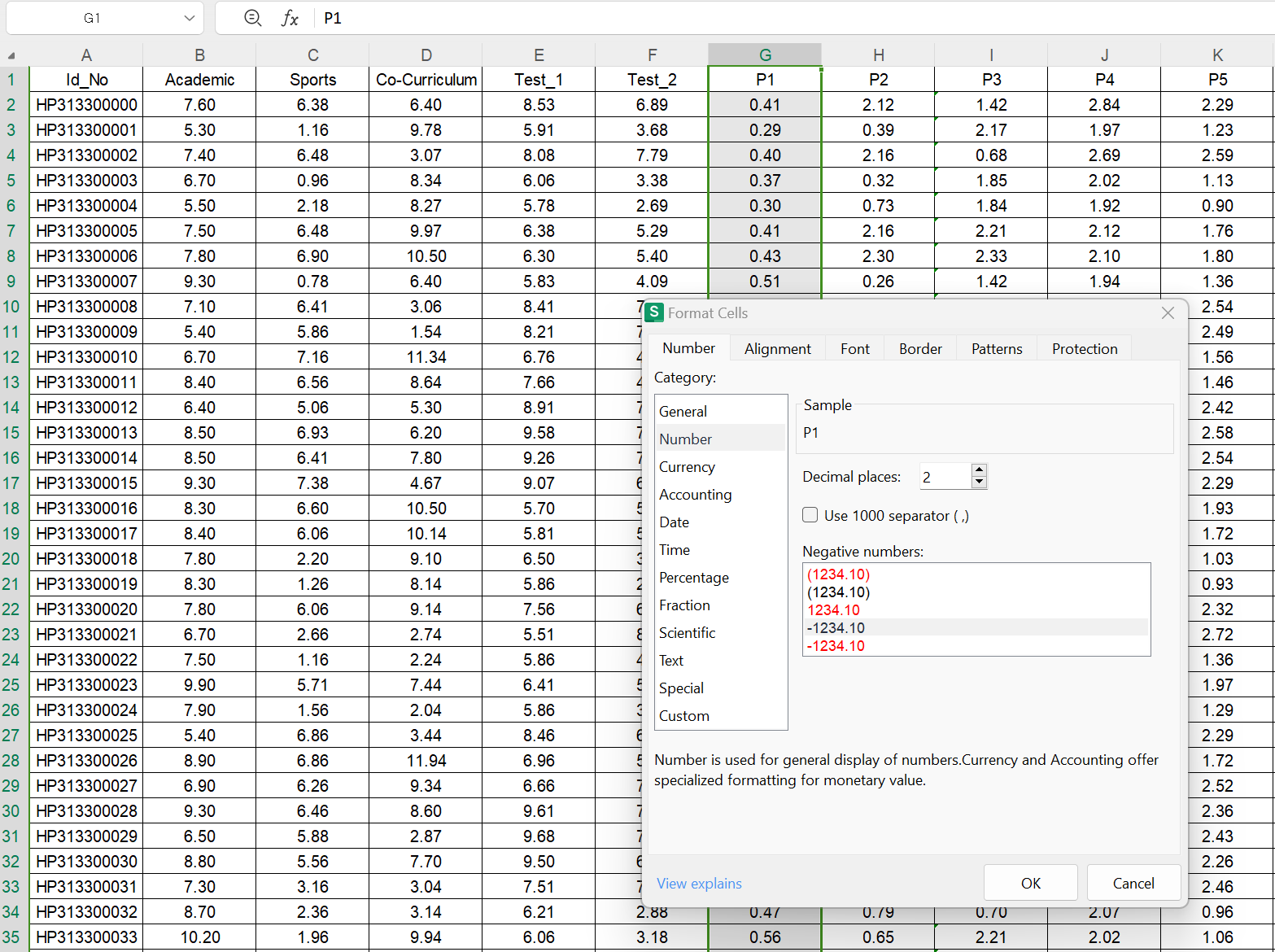
**Introduction:**

The processing and visualisation of a dataset called dataset1.txt is the main topic of the case study. This dataset includes test results of different courses like academic, sports, and co-curriculum. In order to obtain insights into the dataset, the main goals of this case study are to transform, analyse, and develop a graphical dashboard from the data in Google Sheets.

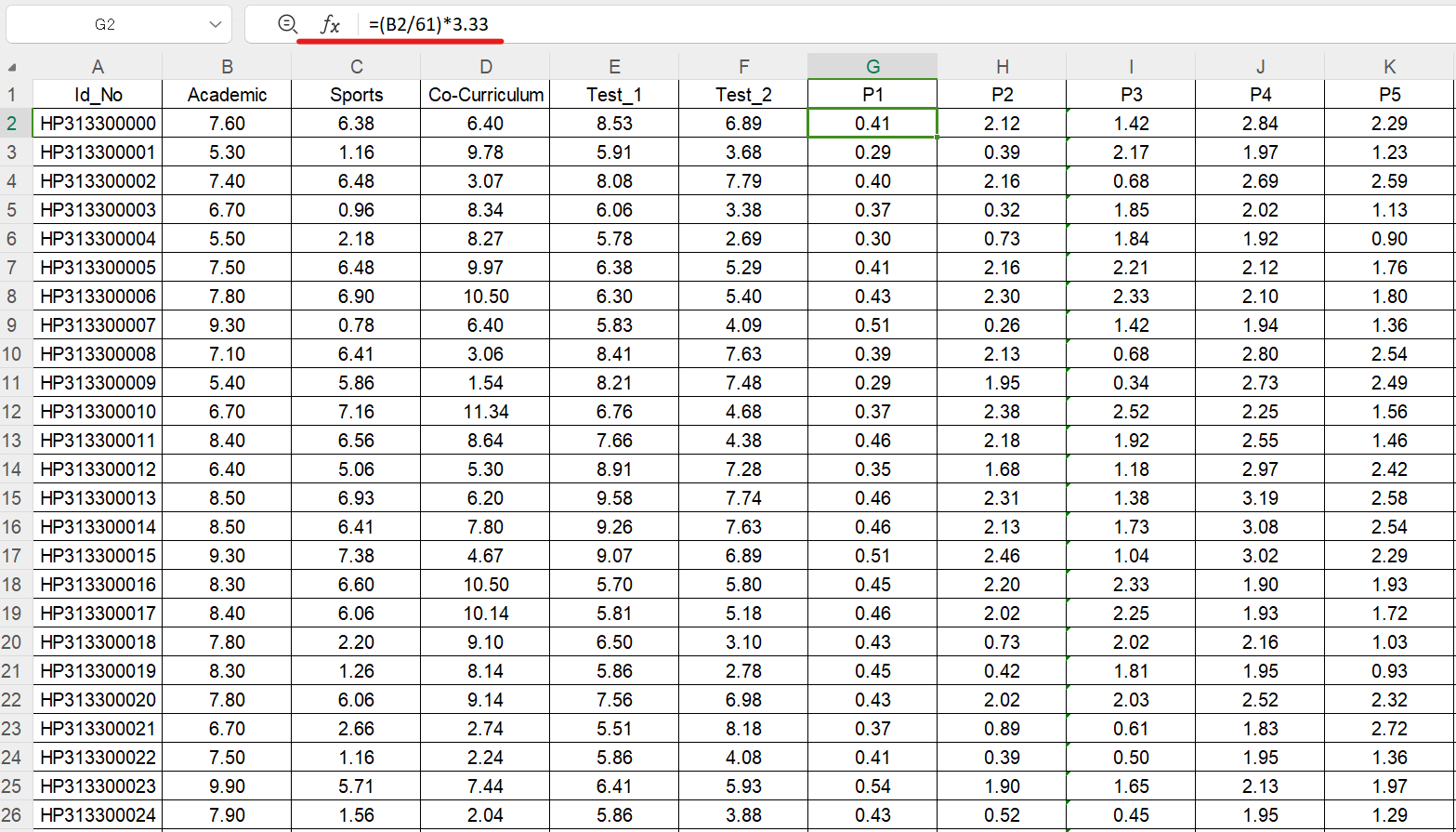
**Data Processing:**

1. A Google Excel sheet was created, and the data from case study 1 was imported there. After that, the Excel sheet link was shared between our team members to work and collaborate together.

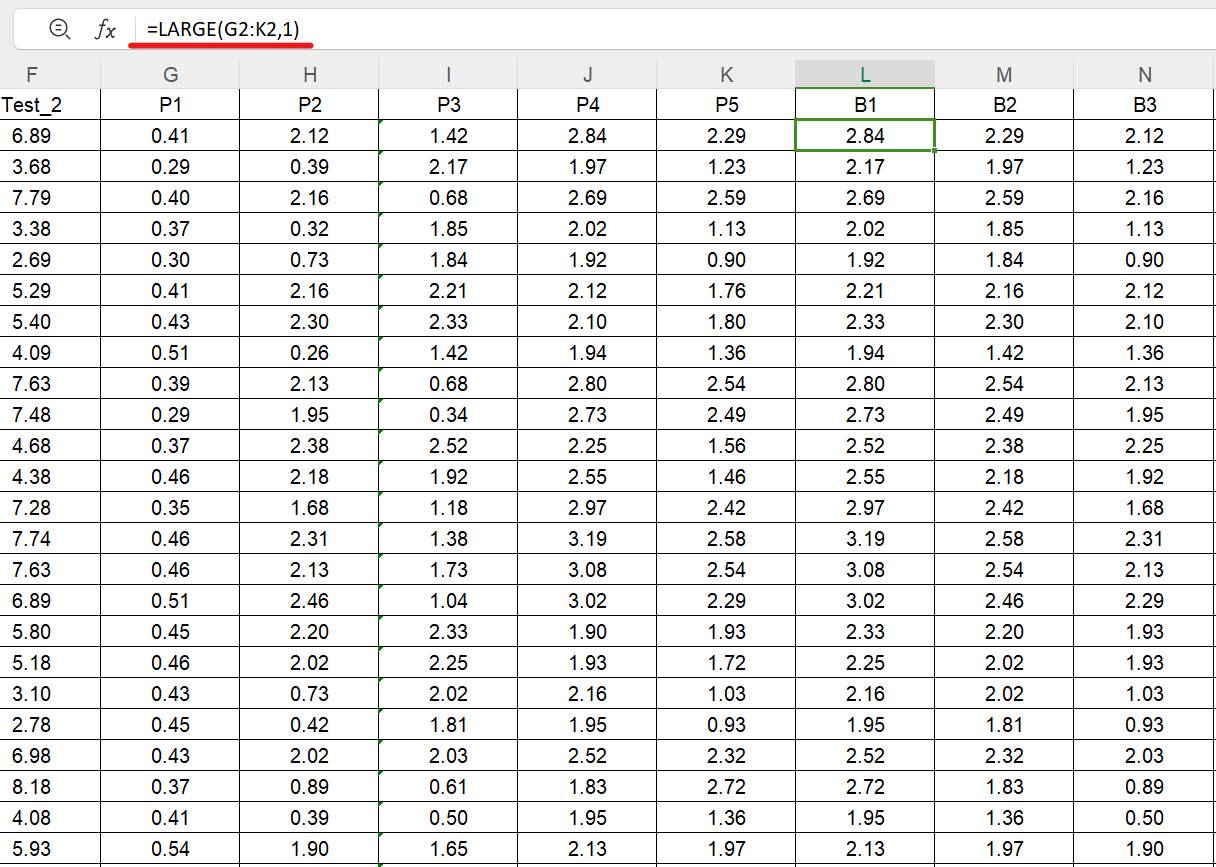
2. To round up the values of the first five columns to 2 decimal places, the excel default round up function from the format cell option has been used. Alternatively, the ROUND function can also be used to do so.



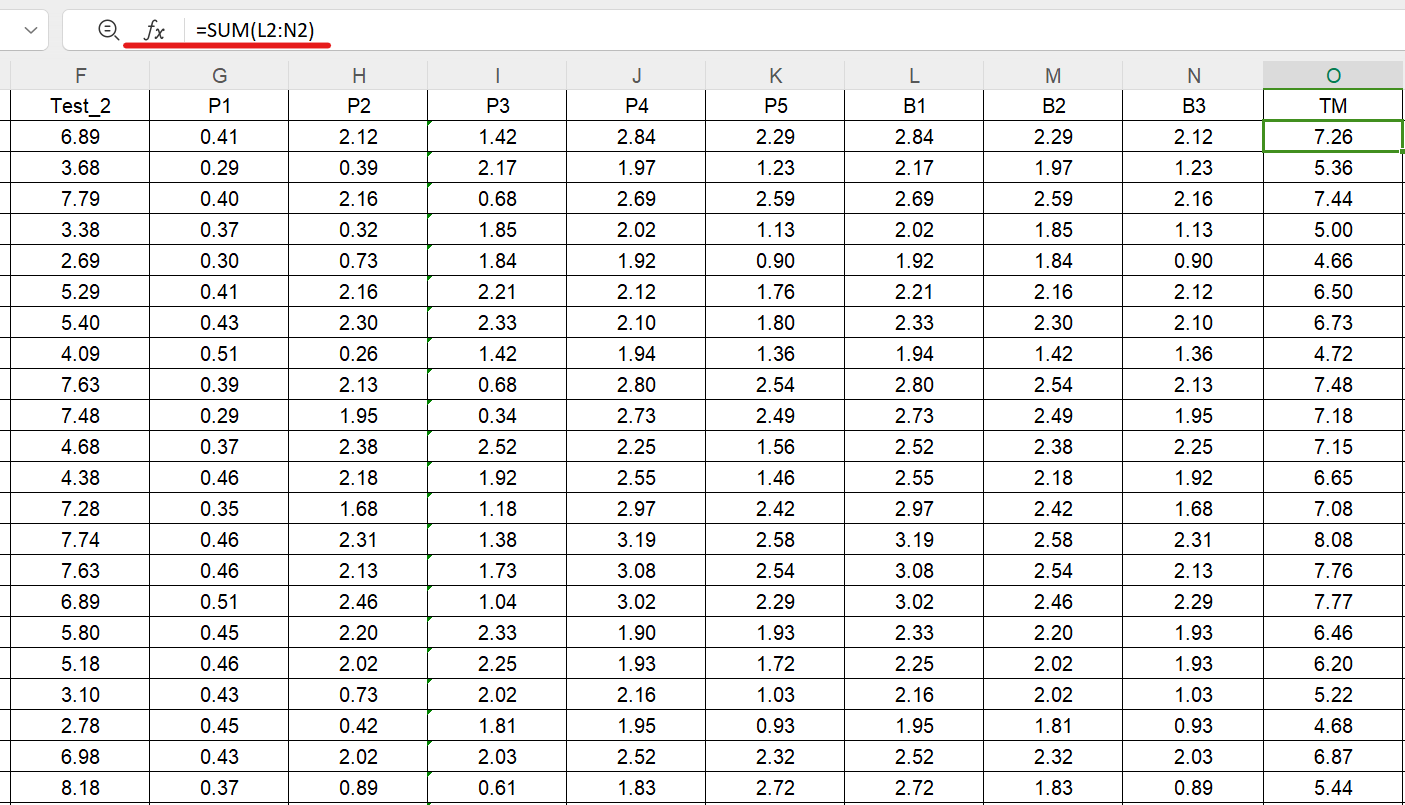
3. To update a new value for columns B (Academic) through F (Test 2), with the new maximum value of 3.33 for each column, at first we have determined the current maximum value in each column. Then, we have used a formula to scale down the values so that the highest value becomes 3.33. The formula would be: =(Original\_Value / Current\_Max\_Value) \* 3.33.



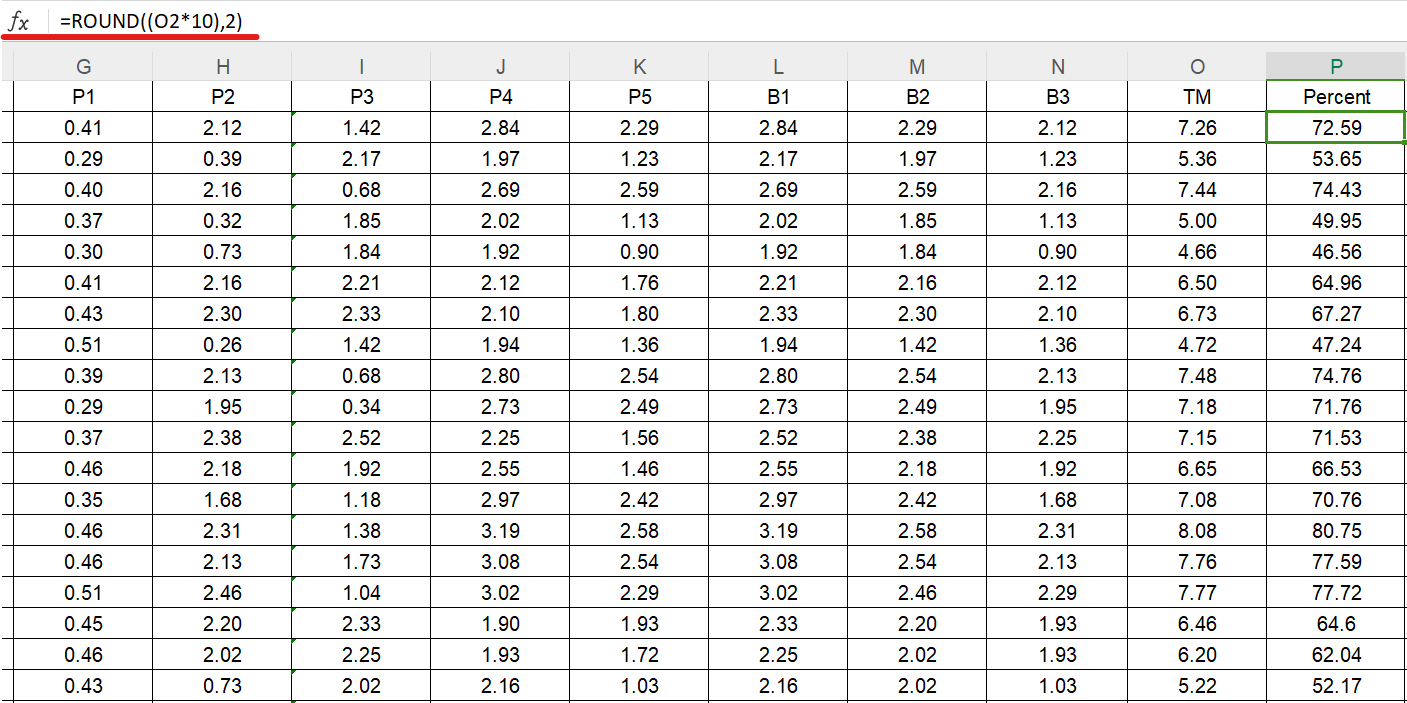
4. To identify the three highest values, the LARGE function is useful. Input =LARGE(G2:G,1) in cell L2 to determine the highest value. For the second highest value, use =LARGE(G2:G,2) in cell M2. Lastly, to ascertain the third highest value, place =LARGE(G2:G,3) in cell N2.



5. To calculate total points from columns L to N. we have used the formula =SUM(L2:N2) in the first cell of column O (TM). Drag this formula down from the first row to apply it to the rest of the column. This will sum and display the total points for each row in column O.

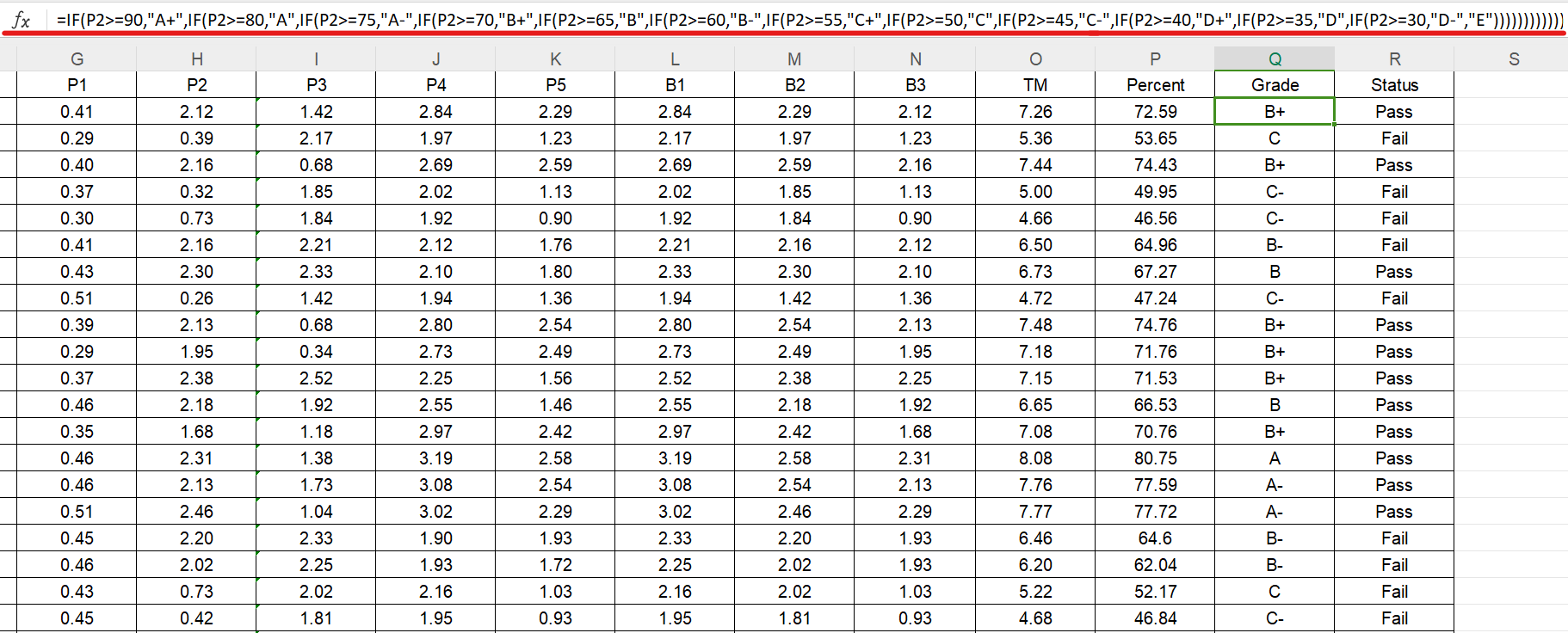


6. To calculate and enter the percentage for data in column O (TM) into column P (Percent) in Excel or Google Sheets, first determine the base value (maximum score) and assume it's in cell X1. In the first cell of column P, enter =ROUND(O2/X1\*100, 2) to calculate the percentage, rounding it to two decimal places. Then, drag this formula down column P to apply it to all rows. This method will display each row's percentage in column P, based on the total marks in column O.



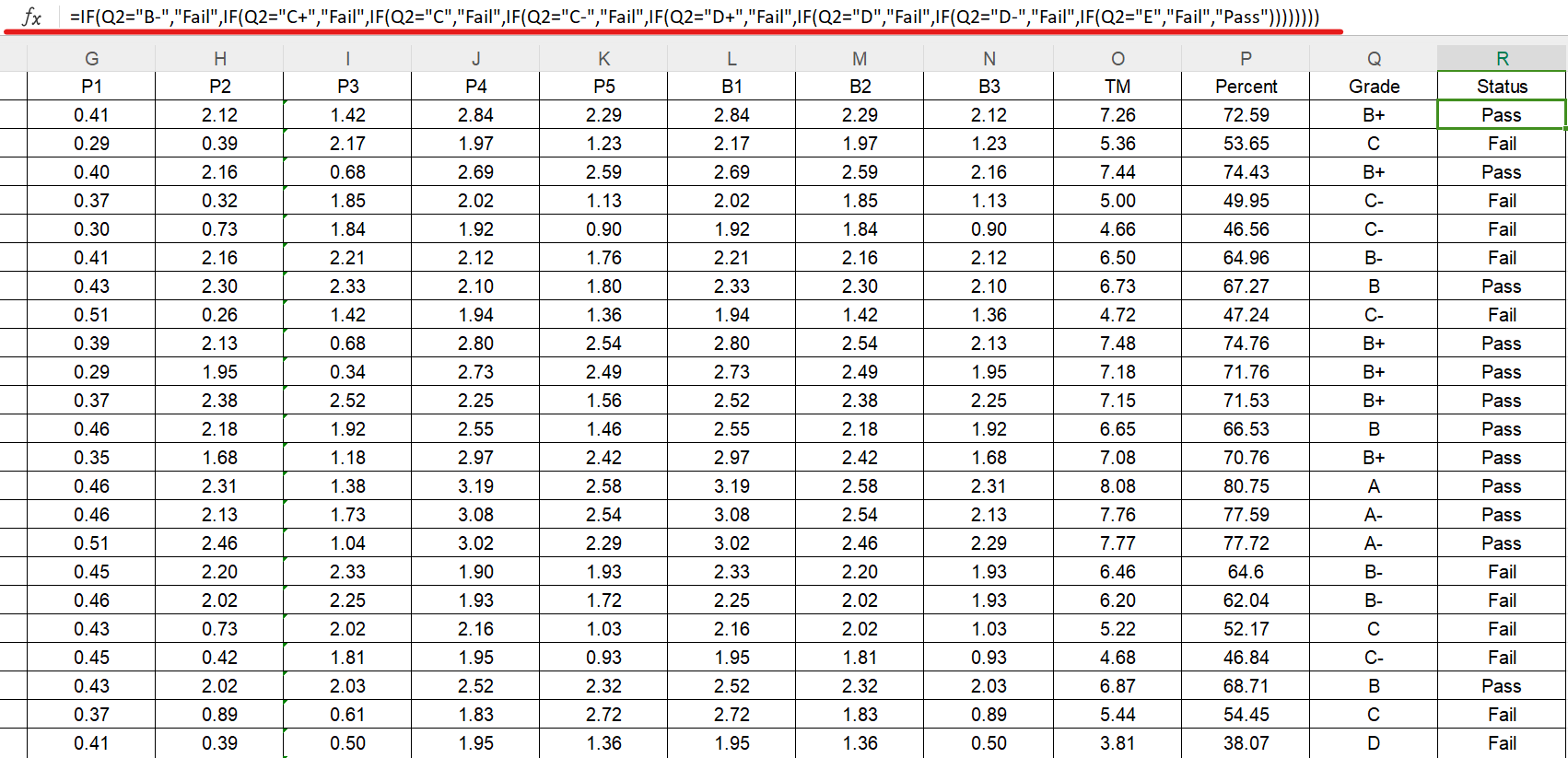
7. Create a New Column for Grade (Column Q): In the first row of column Q (assuming Q2 if you have headers), enter a formula to assign a grade based on the percentage in column P. Use nested IF statements with a grade table to determine the grade. The nested IF formula implemented: is: =IF(P2>=90,"A+",IF(P2>=80,"A",IF(P2>=75,"A-",IF(P2>=70,"B+",IF(P2>=65,"B",IF(P2>=60,"B-",IF(P2>=55,"C+",IF(P2>=50,"C",IF(P2>=45,"C-",IF(P2>=40,"D+",IF(P2>=35,"D",IF(P2>=30,"D-","E"))))))))))))

The formula has been dragged down to apply it to all rows in column Q.



Create a New Column for Status (Column R): In the first row of column R, enter a formula to assign a status based on the grade in column Q. The formula implemented is: =IF(Q2="B-","Fail",IF(Q2="C+","Fail",IF(Q2="C","Fail",IF(Q2="C-","Fail",IF(Q2="D+","Fail",IF(Q2="D","Fail",IF(Q2="D-","Fail",IF(Q2="E","Fail","Pass"))))))))

The formula has been dragged down to apply it to all rows in column R.



Colour Coding in Excel: Excel doesn't allow formula-based cell colouring directly. You'll need to use conditional formatting. Select column P. Go to Conditional Formatting > New Rule. Choose Use a formula to determine which cells to format. Enter a formula like =$R2="Pass" and set the format to green fill. Repeat the process for the "Pass" line with a light red colour, but this time apply the rule to the entire row or the specific cells you want to colour.

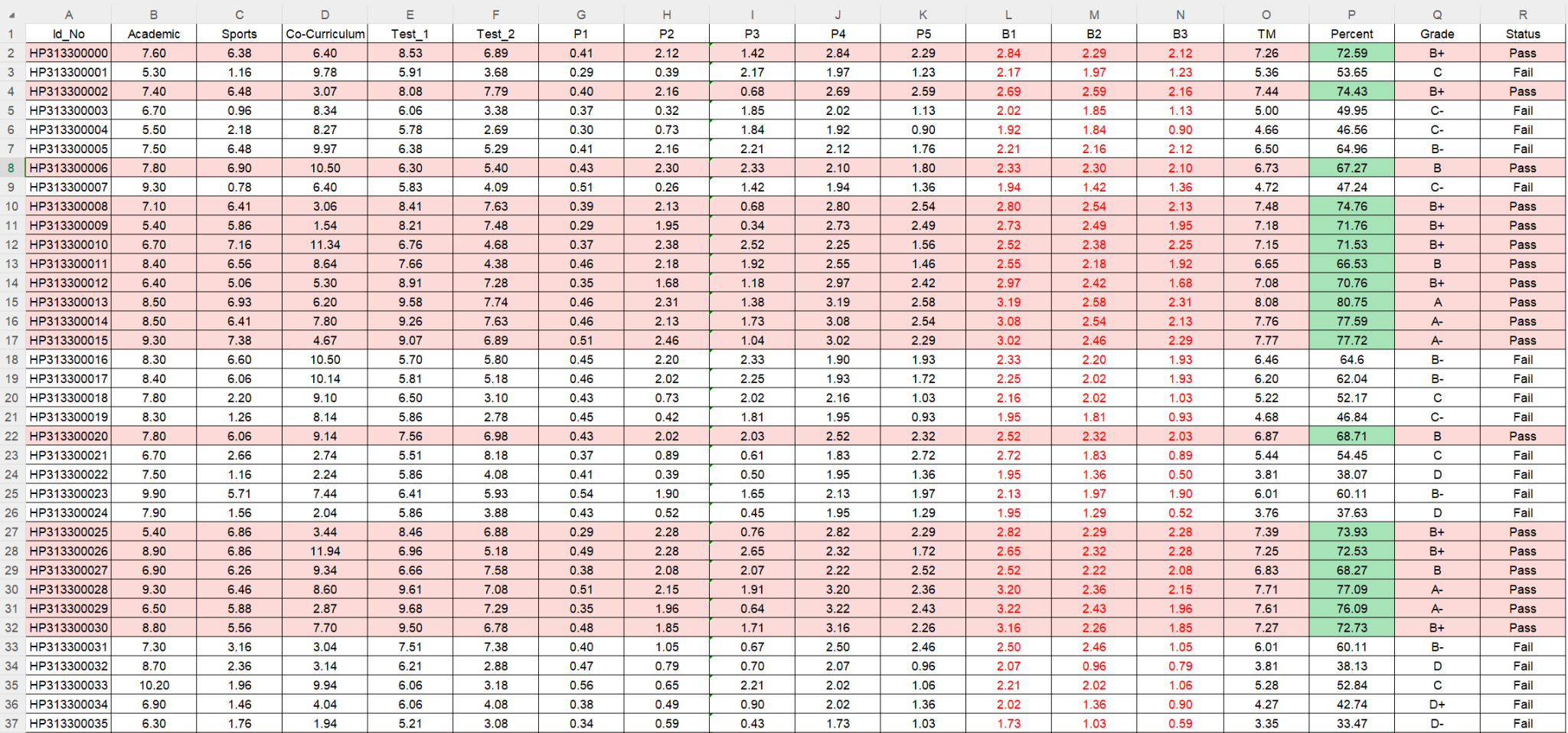


Figure: The final output of the data processing

**Data Visualisation:**

The primary goal of this task is data visualisation, achieved by creating a dashboard in Google Sheets to effectively display key aspects of the data. Atfirst, we created a new sheet and labelled it "Dashboard”. Then we calculated and figured out the following:

* Determining the mean, maximum, and average values
* Displaying the grade results as charts and tables
* The total number of student records for both pass and fail
* Creating pass and fail views in the form of pie charts.

Upon completing the aforementioned steps for data visualisation, the final outcome we observe is as follows:

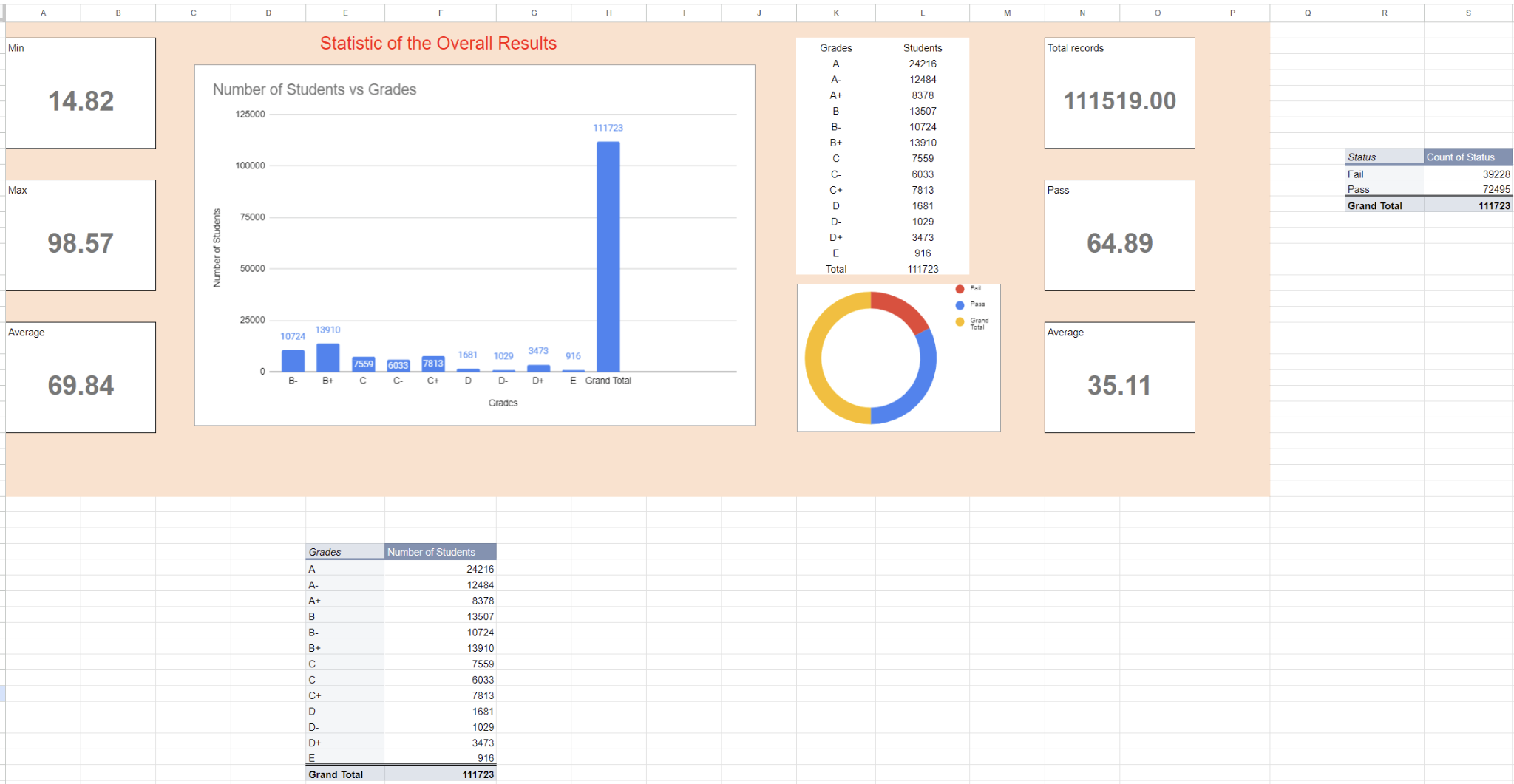


Figure: The final output of the visualisation

**Conclusion:**

To conclude, this assignment effectively showcased data analysis and visualisation in Google Sheets, focusing on student performance across various metrics. Key tasks included data normalisation, point calculation, percentage determination, and grade assignment. The dashboard provided a clear visual representation of the data through graphs, charts, and tables, highlighting important trends like grade distribution and pass/fail ratios. This exercise underscored the value of data analysis skills in education, demonstrating how well-organized data can offer deeper insights and support informed decisions.